Content type: Article

THE SYSTEMATIC CONSERVATION PLANNING FOR INTRASPECIFIC GENETIC DIVERSITY.

Ivan PAZ-VINAS[1,2,3*], Géraldine LOOT[1,4], Virgilio HERMOSO[6], Charlotte VEYSSIERE[1,4], Nicolas POULET[7], Gaël GRENOUILLET[1,4] and Simon BLANCHET[1,5]

[1] Centre National de la Recherche Scientifique (CNRS), Université Paul Sabatier (UPS), École Nationale de Formation Agronomique (ENFA); UMR 5174 EDB (Laboratoire Évolution & Diversité Biologique), 118 route de Narbonne, 31062 Toulouse cedex 4, France.

[2] Aix-Marseille Université, CNRS, IRD, Avignon Université; UMR 7263 IMBE, Équipe EGE, Centre Saint-Charles, Case 36, 3 place Victor Hugo, 13331 Marseille cedex 3, France.

[3] Université de Lyon, CNRS, ENTPE; UMR 5023 LEHNA, Laboratoire d'Écologie des Hydrosystèmes Naturels et Anthropisés, 6 rue Raphaël Dubois, 69622 Villeurbanne, France.

[4] Université de Toulouse, UPS; UMR 5174 EDB, 118 route de Narbonne, 31062 Toulouse cedex 4, France.

[5] CNRS, Station d'Écologie Théorique et Expérimentale; UMR 5321, 09200 Moulis, France.

[6] Centre Tecnologic Forestal de Catalunya, Crta. Sant Llorenc de Monunys, Km 2, 25280, Solsona. Lleida, Spain.

[7] Pôle Écohydraulique ONEMA-IMFT, Office National de l'Eau et des Milieux Aquatiques, France.

Running title: Conservation planning of genetic diversity

Keywords: biodiversity, dendritic ecological networks, fish, freshwater, genetic diversity, geostatistical tools, microsatellites, river basin, spatial patterns, systematic conservation planning

***Corresponding author:**

Ivan PAZ-VINAS
Université Lyon 1
CNRS, ENTPE, UMR 5023 LEHNA
Laboratoire d'Écologie des Hydrosystèmes Naturels et Anthropisés
3, rue Raphaël Dubois - Bât. Darwin C
69622 Villeurbanne Cedex France
Phone: (+33) 04 72 43 26 03
E-mail: ivanpaz23@gmail.com

Co-author's e-mails:

Simon Blanchet: simon.blanchet@sete.cnrs.fr
Virgilio Hermoso: virgilio.hermoso@gmail.com
Charlotte Veyssière: veyssiere.charlotte@gmail.com
Nicolas Poulet: nicolas.poulet@onema.fr
Gaël Grenouillet: gael.grenouillet@univ-tlse3.fr
Géraldine Loot: geraldine.loot@univ-tlse3.fr

Introductory paragraph

Intraspecific diversity is a fundamental facet of biodiversity allowing species to adapt to environmental changes. Preserving intraspecific diversity should hence be a main conservation target, although attempts to identify priority conservation areas for intraspecific diversity remain scarce. Using molecular data on six freshwater fish species sampled at a large spatial scale, we determined hot- and cold-spots of genetic diversity, and identified priority conservation areas using a systematic conservation planning approach. We demonstrated that the systematic conservation planning of intraspecific diversity is feasible. However, we found poor levels of congruency and surrogacy among conservation solutions found for each species, indicating that the conservation of intraspecific genetic diversity should be species-specific. This was because of the strong among-species incongruences in hot- and cold-spots of genetic diversity. Our study demonstrates the usefulness of systematic conservation planning for preserving intraspecific diversity, and it provides guidelines to identify priority conservation areas for intraspecific genetic diversity.

Biodiversity conservation is a major challenge that is often addressed by identifying protected areas with high biodiversity and/or landscape values[1]. Conservation areas are generally identified as areas with high proportions of endemic, rare or iconic species[2]. Alternatively, conservation planning can be based on the concept of complementarity between conservation areas[3], and on cost-effectiveness analyses such as systematic conservation planning methods (hereafter SCP[4]). SCP aims at identifying a number of sites (i.e. *irreplaceable sites* that should be managed for conservation in priority) best representing a predefined proportion of the biodiversity observed in a region, at a minimum cost.

There have been attempts to include information on the phylogenetic history of species assemblages into SCP approaches to preserve both species identities and their macro-evolutionary history[5,6]. However, genetic diversity observed at the *population*

level (i.e. within species) has rarely been considered in SCP (but see[7–9]). *Intraspecific genetic diversity* is a fundamental facet of biodiversity, as it is the fuel for species to adapt to global and environmental changes[10–13]. Conservation geneticists have classically considered this facet of biodiversity in conservation plans, for instance by identifying "Evolutionary Significant Units" or unique genetic lineages[14]. However, conservation geneticists have yet ignored the possibility to combine genetic data (e.g. allele identities) to dedicated planning tools such as SPC (but see[7,8]).

The relative lack of genetic datasets at *large* spatial scales may partly explain why SCP has yet rarely been applied to intraspecific genetic diversity[15,16]. Particularly, conservation geneticists have been generally restricted by the amount and spatial range of datasets (but see[17–19]). However, our capacity to compile genetic datasets at large spatial, temporal and taxonomic scales has greatly increased in the last decades[20,21], so that it is now possible to identify priority areas for the conservation of genetic diversity using dedicated conservation planning tools.

Here, we tested the potential of SCP analyses to identify priority conservation areas accounting for intraspecific genetic diversity measured at a large spatial scale. We first considered a set of four common and representative freshwater fish species (*Squalius cephalus*, *Gobio occitaniae*, *Barbatula barbatula* and *Phoxinus phoxinus*) to test the influence of conservation targets (proportion of the total amount of genetic diversity to be covered by irreplaceable sites) and analytical strategies (analysing each species independently or all species pooled) on final conservation solutions (number and identity of irreplaceable sites). We then included two rare species of particular conservation interest (*Leuciscus burdigalensis* and *Parachondrostoma toxostoma*) to test the relevance of the SCP approach in a "real conservation-oriented study". For these two species (and the four common species), we ran SCP analyses considering a typical

conservation target[22,23] to (i) explore the spatial distribution of irreplaceable sites in a riverscape, and (ii) test for congruency and surrogacy in irreplaceable sites among species, and more particularly between rare and common species. We finally tested whether –or not– irreplaceable sites were correctly predicted by classical indices of genetic diversity (e.g. allelic richness). We demonstrate that preserving the genetic diversity of a species assemblage is a feasible –yet complex– task necessitating appropriate analyses to assist the decision-making process.

Results

Descriptive statistics. We assessed for the six species within-sites intraspecific genetic diversity (i.e. α-IGD) by calculating both allelic richness (*AR*) and richness in private alleles (*PA*). Overall, *P. toxostoma* (one of the two rare species) showed the lowest α-IGD. Mean *AR* ranged from 2.114 for *P. toxostoma* to 5.821 for *P. phoxinus*, and mean *PA* ranged from 0.036 for *P. toxostoma* to 0.162 for *L. burdigalensis* (Table S1). We also assessed among-sites intraspecific genetic diversity (i.e. β-IGD) by quantifying (for each species) how much a site is genetically unique compared to all others (using the D_{est} index[24], see the Methods section). *Parachondrostoma toxostoma* also showed the lowest mean D_{est} value (0.069), while the highest mean value was found for *B. barbatula* (0.383; Table S1).

<u>Testing the suitability of SCP analyses for intraspecific genetic diversity</u>

In this first step we focused on the four common species for which genetic data were available at a large spatial scale, allowing a thorough exploration of the suitability of SCP for intraspecific genetic diversity.

Spatial patterns of genetic diversity. Using Generalized Linear Models for Stream Networks (GLMSSN)[25,26], we showed that spatial patterns of genetic diversity largely varied and were actually poorly congruent among the four common fish species (Figure 1). As extreme examples, hotspots of *AR* for *S. cephalus* where mainly found in the Western part of the network and on the core streams, whereas these same areas were identified as coldspots of *AR* for *B. barbatula* (Figure 1A1-1A3). Similarly, hotspots of *PA* were inversely related between *G. occitaniae* and *P. phoxinus* (Figure 1B2-1B4). Similar conclusions were reached for D_{est} (Figure 1C). For instance, hotspots of D_{est} were observed in opposite areas of the river basin for the species pair *B. barbatula/P. phoxinus* (Figure 1C3-1C4). As a consequence, the sign, slope and significance of GLMSSNs explanatory variables strongly varied among species (Figure 1). This incongruence in spatial patterns of genetic diversity among species was also reflected by the low to moderate correlation coefficients measured among all possible pairs of species and for each index of genetic diversity (i.e. *AR*, *PA* and D_{est}, Table S2). Indeed, Pearson's correlation coefficients were lower than 0.6 for all comparisons but two (i.e. between *B. barbatula*/*G. occitaniae* and between *P. phoxinus*/*B. barbatula* for *AR*, see Table S2).

The influence of conservation targets and analytical strategies to identify irreplaceable sites for genetic conservation. We used alleles' presence/absence data combined to SCP procedures to test the influence of conservation targets and analytical strategies on the

number and identity of irreplaceable sites (i.e. sites that were selected in all the solutions in SCP analyses; see the Methods section for more details).

When species were analysed independently, we found that the number of irreplaceable sites increased as the conservation target (i.e. the percentage of alleles to be covered by irreplaceable sites) increased, with a steep increase for conservation targets higher than 75% of the total number of alleles present at the river basin scale (Figure 2). However, the percentage of irreplaceable sites strongly varied among species. For instance, for a 90% conservation target –which is a common target in conservation genetics[22]–, the proportion of irreplaceable sites ranged from 3.61% of the total number of sampled sites for *G. occitaniae* to 28.57% for *P. phoxinus* (Table S3; Figure 2). For extreme conservation targets (100% of alleles to be covered), the proportion of irreplaceable sites varied from 25.30% for *G. occitaniae* to 68.26% for *P. phoxinus* (Table S3; Figure 2).

When alleles from the four common species were analysed in a single pooled analysis, we similarly found that the proportion of irreplaceable sites increase as the conservation target increases (Figure 2). Interestingly, we did not identify irreplaceable sites for the 30% conservation goal, and only 3 irreplaceable sites were found for the 50% conservation goal (Figure 2). The proportion of irreplaceable sites increased moderately to 17.39% for the 75% target, and then steeply increased for higher conservation targets to reach 55.70% for the 90% target and 76.08% for the 100% target (Figure 2). This later result suggests that almost all the river basin should be protected to reach high conservation targets when adopting a pooled strategy.

<u>A real conservation-oriented study using SCP approaches</u>

We here focused on two rare species (in addition to the four common species) to explore the usefulness of SCPs in a real case study.

Identification of irreplaceable sites for genetic conservation. We first visually explored the spatial distribution of irreplaceable sites. Overall, the localization of irreplaceable sites in the riverscape strongly varied among species, being spread all over the river basin (Figure 3). We failed to identify areas (e.g. upstream or downstream locations) clustering irreplaceable sites for any species (Figure 3A-F). This apparent lack of clustering was statistically confirmed by our generalized linear models (Table S4). Indeed, the two positional indices we used as explanatory variables (i.e. distance from the outlet of sampling sites and the betweenness centrality of each sampling site, see the Methods section) were not significant predictors of the irreplaceability of sites for all species, except for distance from the outlet for *P. phoxinus* (Table S4). This indicates that neither the position of sites in the riverscape, nor the positional importance of these sites, determine the irreplaceability of sites.

Second, we tested whether conservation solutions found for each species were spatially congruent among species. Over the six fish species, we identified forty-two sites (out of the ninety-two sites, i.e. 45.65%) as irreplaceable at the 90% conservation target for at least one species (Figure 4). Thirty-two of these forty-two sites were irreplaceable for at least one of the four common species (Figure 4), and fourteen of the forty-two sites were irreplaceable for at least one of the two rare species (Figure 4). Among the six species, only eight out of these forty-two sites were irreplaceable for at least two species, and only one of these sites was irreplaceable for three species (Figure 4).

Surrogacy in irreplaceable sites among species. We tested whether solutions found for one species can be used as a surrogate for other species by calculating the percentage of the total number of alleles observed for a given species that is covered by irreplaceable sites identified for another species. Levels of surrogacy were generally low to moderate, and strongly varied among species pairs (Table 1). For instance, irreplaceable sites identified for *G. occitaniae* failed to cover the genetic diversity of the two rare species, and covered only 17.66 to 42.60% of the total number of alleles of the other common species. This indicates that irreplaceable sites identified for *G. occitaniae* (i.e. the most widespread species) are poor surrogates for preserving the intraspecific genetic diversity of other species (18.38% of surrogacy; Table 1). Conversely, irreplaceable sites found for *P. phoxinus* and *L. burdigalensis* are better surrogates for *G. occitaniae*, as 79.58% and 70.27% of the total number of alleles of *G. occitaniae* was covered by irreplaceable sites identified for *P. phoxinus* and *L. burdigalensis* respectively (Table 1). Overall, irreplaceable sites best covering genetic diversity of other species were those identified for *B. barbatula*, which covered in average 68.28% of the total number of alleles of other species (Table 1).

Interestingly, the thirty-two irreplaceable sites identified for the four common species covered 79.87% and 90% of the total number of alleles of *L. burdigalensis* and *P. toxostoma* respectively, suggesting that irreplaceable sites identified for a set of common species can be good surrogates for intraspecific genetic diversity of rare species. Conversely, the fourteen irreplaceable sites identified for the two rare species covered a total number of alleles ranging from 65.18% for *S. cephalus* to 79.40% for *B. barbatula*.

Relationships between genetic indices and irreplaceable sites. For all species but *G. occitaniae* and *P. toxostoma* (for which none of the variables were significant predictors), *PA* was identified as the only variable that significantly predicted the probability for a site to be identified as an irreplaceable site (Table S5). This probability increased with the number of *PA* in a site.

Discussion

Intraspecific diversity constitutes the fuel for species and populations to cope with environmental changes[10–12]. This biodiversity facet is hence the first that should respond to global change, allowing populations and species to respond adaptively to these changes[13]. However, this biodiversity facet has so far been poorly integrated in dedicated optimization planning tools (but see[7,8]). We here fill this gap by demonstrating that the systematic conservation planning of intraspecific genetic diversity is a feasible –yet complex– task necessitating careful considerations.

From idiosyncratic distributions of genetic diversity…

Our results strongly suggest that the genetic diversity of the targeted species did not follow a common spatial pattern, but rather species-specific (idiosyncratic) spatial distributions. This conclusion holds true for all genetic diversity indices, and it corroborates the few previous studies investigating simultaneously the spatial distribution of genetic diversity at large spatial scales and for sympatric species (e.g.[17,27]). This was however unexpected given that recent meta-analyses on freshwater organisms demonstrated that α-IGD is generally higher in downstream than in upstream areas[28], and that β-IGD tends to be higher in upstream than in downstream sections[29].

Overall, these very general spatial patterns of intraspecific genetic diversity were verified in our datasets. For instance, a negative relationship between allelic richness and distance from the outlet is expected in freshwater organisms[28], and was actually observed for three out of the six species considered in this study (Figure S1). However, when using a precise and novel approach to map genetic diversity across the network, we demonstrated that the distribution of cold- and hot-spots of α- and β-IGD was subtler and idiosyncratic. This probably reflects interactions between colonization histories, life-history-traits of species and the network structure, which are expected to drive patterns of genetic diversity in rivers[28,30,31].

…to the systematic conservation planning of intraspecific genetic diversity.

The spatial mismatch in intraspecific genetic diversity among species probably explains why we found that the level of congruency and surrogacy of irreplaceable sites identified for each species was extremely low. For instance, there was an extremely low proportion of irreplaceable sites that were common to two or three species (and never more than three species; Figure 4). In the same way, we detected no clear patterns in the spatial distribution of irreplaceable sites. In riverscapes, it is generally assumed that small upstream areas are the "source" of genetic uniqueness, and hence the primary areas to protect (i.e. the "small but mighty" paradigm[29]). Our results did not confirm this paradigm since irreplaceable sites (for any of the six species) were not particularly situated in upstream areas and/or in areas of high connectivity (i.e. areas displaying high centrality values[18,32]), and rather suggests that priority areas for the conservation of intraspecific genetic diversity should cover the whole distribution range of species. Finally, the level of surrogacy among irreplaceable sites was low to moderate, and never attained the 90% threshold we assumed when we considered all species

independently[22,23]. Combined to our finding that the number of irreplaceable sites can be relatively high for reasonable conservation targets (up to 46% sites are identified as irreplaceable sites for at least one species at the conservation target of 90%), we concluded that our ability to identify priority areas for intraspecific genetic diversity is highly species-specific and depends on the capacity to tackle the trade-off between the amount of genetic diversity to protect, and the extent of priority areas we can realistically protect.

However, when surrogacy between all irreplaceable sites identified for the entire set of common species and those identified for the rare species was tested, we reached reasonable proportions of the total number of alleles to protect (~80-90%) for the two rare species (i.e. *P. toxostoma* and *L. burdigalensis*). This result suggests that, in some cases, genetic data obtained for a set of widely-distributed, "easier-to-sample" common species displaying varying life-history traits can be used for identifying protection areas for the intraspecific genetic diversity of other sympatric rare species that can be more problematic to sample.

Overall, our results suggest that two different analytical strategies can be employed in real-case SCP studies aiming at preserving intraspecific genetic diversity (i) identification of conservation areas for each rare species independently or (ii) identification of conservation areas for a set of representative common species. Both strategies have their own advantages and inconveniences. The first strategy optimally preserves genetic diversity of rare species at competitive costs (e.g. 14 irreplaceable sites to protect in our demonstrative study), but this at the expense of the genetic diversity of other sympatric species. Conversely, the second strategy will optimally preserve genetic diversity of a set of common species while maintaining high levels of genetic diversity for rare species, but this at a higher cost (e.g. 32 irreplaceable sites to

protect in our study). Whether to choose one of these two strategies will therefore depend on many factors such as how difficult is the sampling of rare species compared to common species, or the extent of the resources available for setting new protected areas. We recommend however to adopt the second strategy when possible, since it allows to simultaneously maintain genetic diversity from rare and common species. Indeed, genetic diversity of common species is vital for ensuring ecosystem stability, as it ultimately influence species interactions, population dynamics and ecosystem functions[33], and we argue that it should be considered in conservation plans.

Conclusions

Our study provides novel, insightful and promising knowledge on the setting of priority conservation areas for intraspecific diversity. It shows that systematic conservation planning methods are useful objective tools for conservation geneticists whose conservation solutions will strikingly depend on the species to be preserved and the quantity of genetic information that managers aim at preserving in a landscape. Given our results, we suggest that two strategies could be employed in real-case conservation programs: (i) identification of priority conservation areas for each rare species independently or (ii) identification of priority conservation areas on the basis of the analysis of a set of representative common species that may serve as “umbrellas” for rare sympatric species.

Our study also raises many additional questions that should be considered in the near future. Among others, we believe that the next steps will be to formally identify sound conservation targets for intraspecific diversity, to test whether neutral intraspecific diversity appropriately mirrors quantitative and adaptive diversity[12], and to

quantify the influence of intraspecific diversity on ecosystem functioning and services, so as to better evaluate the added value of preserving such a facet of biodiversity[12,34].

Methods

Data collection

Biological models. We focused on an assemblage of six Cyprinid freshwater fish species. Three of them are widespread in Europe (i.e. *Squalius cephalus*, *Phoxinus phoxinus* and *Barbatula barbatula*) whereas three of them (*Gobio occitaniae*, *Leuciscus burdigalensis* and *Parachondrostoma toxostoma*) are endemic to Southern France[35]. This set of species covers a large functional trait space that is representative of many freshwater fish communities. For instance, *S. cephalus* is a large-bodied fish species with long lifespan (i.e. it can be 60 cm long and live up to 15 years[35]) whereas at the extreme *P. phoxinus* is a small-bodied species with shorter lifespan (i.e. it is less than 12 cm long and usually lives up to 4-5 years[35]). From an ecological perspective, *G. occitaniae*, *P. toxostoma* and *B. barbatula* are bottom feeders, whereas *S. cephalus* and *P. phoxinus* are water column feeders and *L. burdigalensis* is more opportunistic. Further, *B. barbatula* is mainly active during night, while the other species are particularly active during the day. Four of these species are relatively abundant (i.e. *S. cephalus*, *G. occitaniae*, *P. phoxinus* and *B. barbatula*), although they greatly vary in their ecological niche and hence their spatial occupancy in the river network (see Figure S2 for maps representing the spatial distribution of sampling sites for each species, which roughly corresponds to their spatial distribution in the Garonne-Dordogne river basin). We will hereafter refer to this set of species as the “common” species. The two

other species are rare (*L. burdigalensis*) to very rare (*P. toxostoma;* see Figure S2) in the Garonne-Dordogne river basin, and are of particular interest for conservation. *Leuciscus burdigalensis* is a recently described species that is locally experiencing both demographic and genetic bottlenecks in many populations[36,37]. *Parachondrostoma toxostoma* is a vulnerable species[38] listed in the IUCN red list, in the Annex II of the European Union Habitats Directive and in Appendix III of the Bern Convention[38].

Sampling design. During Spring/Summer 2010-2011, we used electric-fishing to sample ninety-two sites distributed across thirty-five rivers from a large river basin, the Garonne-Dordogne River basin (>100,000 km^2, South-Western France; Figure S2; Table S6). Sampling sites were chosen to cover the whole distribution range of each species at the riverscape scale, and to allow characterising spatial patterns of genetic diversity for all these species. Up to 25 individuals *per* species *per* site were sampled when possible. Not all species were present at all sampling sites (Figure S2; Table S6), and some species were at a density that did not allow reaching the 25 individuals threshold. In these cases, we captured as many individuals as possible. We anesthetized each individual and then we collected and stored in 90% ethanol a fragment of their pelvic fin. All individuals were released alive at their sampling location.

Genotyping. Genomic DNA was extracted using a salt-extraction protocol[39]. We used multiplexed Polymerase Chain Reactions (PCRs) to co-amplify 8 to 15 microsatellite loci depending on the species (8 for *G. occitaniae*, 9 for *B. barbatula*, 10 for *S. cephalus* and *P. phoxinus*, 14 for *L. burdigalensis* and 15 for *P. toxostoma*). We used 5-20 ng of genomic DNA and QIAGEN® Multiplex PCR Kits (Qiagen, Valencia, CA, USA) to perform PCR amplifications. Details on loci, primer concentrations, PCR conditions and

multiplex sets can be found in Table S7. The genotyping was conducted on an ABI PRISM™ 3730 Automated Capillary Sequencer (Applied Biosystems, Foster City, CA, USA). The scoring of allele sizes was done using GENEMAPPER® v.4.0 (Applied Biosystems).

Genetic diversity assessment. Given the good spatial resolution of the sampling obtained for the four common species (i.e. *S. cephalus*, *G. occitaniae*, *P. phoxinus* and *B. barbatula*), descriptive genetic analyses were conducted for sampling sites displaying a minimum sample size of N=10 individuals for these species, so as to maximize consistency on subsequent allelic frequency-based genetic analyses (see Figure S2 and Table S6 for details on sample sizes). For the two rare species, for which the sampling was more restricted (i.e. *L. burdigalensis* and *P. toxostoma*), genetic analyses were conducted for sampling sites displaying a minimum sample size of N=6 individuals, so as to maximize the number of sampling sites included in the SCP procedures. We then determined for each of the six species the occurrence of null alleles and potential scoring errors with the program MICROCHECKER 2.3[40]. We tested for departures from Hardy-Weinberg (HW) equilibrium with the 'adegenet' R package v1.6-2[41]. The program GENEPOP v4.0[42] was used to assess linkage disequilibrium among loci within sites. We found significant deviations from HW for a few locus/population pairs for each six species (see Appendix A1 and Supplementary File 1 for details and raw tables), and significant linkage disequilibrium and homozygote excesses for only the four common species (Appendix A1; Supplementary File 1). However, no clear patterns were observed for any species across loci and populations for these deviations. Given the small extent of these deviations and given the large spatial extent of the databases,

we assumed that they weakly affected our main findings (Appendix A1; Supplementary File 1).

To assess within-sites intraspecific genetic diversity (i.e. α-IGD), we applied rarefaction procedures implemented in ADZE v1.0[43] to calculate both allelic richness (*AR*[44]) and private allelic richness (*PA*[45]) at the sampling site level (based on a minimum of N=10 individuals for common species or N=6 individuals for rare species). To assess the among-sites component of intraspecific genetic diversity (i.e. β-IGD), we used the R package 'mmod'[46] to calculate –for each species– a pairwise genetic differentiation index (i.e. D_{est}[24]). For each site (and species), we then derived the averaged value of all pairwise D_{est} values estimated between one given site and all the remaining sites, so as to obtain a single value *per* site.

<u>Testing the suitability of SCP analyses for intraspecific genetic diversity</u>

In the first step, we tested the influence of conservation targets and analytical strategies on final conservation solutions. In this step, we focused specifically on data from the four common species, as their large coverage of the sampling area is more suited for the demonstrative exercise done in this step.

Spatial patterns of genetic diversity. We first used geostatistical modelling tools to explore spatial patterns of α and β genetic diversity for the four common species at the riverscape scale by predicting the distributions of *AR*, *PA* and D_{est} from the observed empirical values using Generalized Linear Models for Spatial Stream Networks (GLMSSN[25,26]). This was done using the 'STARS' toolset of ARCGIS v10.2 and the R package 'SSN'[25,26]. We conducted a model selection procedure based on a comparison

of Akaike Information Criteria (AIC) estimated for several competing GLMSSNs. These models were built by (i) assuming three geographic descriptors (i.e. topological distance from the outlet, longitude and latitude) as explanatory variables, and (ii) choosing a tail-down covariance structure model among the following ones: exponential, Mariah, spherical, linear-with-sill and Epanechnikov. As the number of explanatory variables differed among the GLMSSNs we built, we used the maximum likelihood estimation method for each GLMSSN, so as to allow AIC-based model comparisons. For each common species and genetic index, the best model had the lowest AIC score (see Supplementary File 1 for raw results). This model was used to estimate the slope and the significance of the relationships between explanatory variables and each genetic index, so as to test whether or not spatial patterns of intraspecific genetic diversity can be detected. We finally used predictions from the best models to produce krigged maps for each common species and each genetic index, so as to visually represent the spatial distribution of intraspecific genetic diversity across the whole river drainage, and to visually highlight hot- and cold-spots of intraspecific genetic diversity. We also calculated Pearson's correlation coefficients between values calculated at the site level for each genetic index (i.e. *AR*, *PA* and D_{est}) for each pair of common species, so as to test for spatial congruency in patterns of genetic diversity among common species.

Identification of irreplaceable sites. We then tested whether conservation targets (i.e. the percentage of total number of conservation units to be present in the final conservation solution) and analytical approaches (i.e. species-specific or species-pooled analyses) influence the identification of irreplaceable sites. SCP methods traditionally use species presence/absence data as input data to identify irreplaceable sites for the

conservation of taxonomic diversity at the community level (i.e. sites that cannot be excluded from an optimal selection of sites for conservation)[47]. Here, we replaced species presence/absence data by alleles' presence-absence data to identify irreplaceable sites for the conservation of intraspecific genetic diversity of each species in the river drainage. We used the program Marxan v2.1[47] and genetic data from the common species to identify, for each common species independently, an optimal set of sites that best represent at least 50, 75, 90 or 100% of the total number of alleles present in the whole riverscape at a minimum "cost", which corresponds to four conservation targets. Given the lack of ground estimates for conservation cost, we used a constant cost *per* site, so our objective translated into identifying the minimum number of sites that represent a given proportion of intraspecific genetic diversity (i.e. 50, 75, 90 or 100% of the total pool of alleles represented at least once[48]). We arbitrarily choose the 50, 75 and 100% conservation targets to explore how the proportion of alleles to protect affects the selection of irreplaceable sites. We additionally tested the 90% conservation target, as it corresponds to a threshold target being typically assumed in *ex-situ* conservation plans[22,23]. In order to estimate the relevance of each site to preserve a given proportion of the allelic diversity in the river basin, we used two different methods. In case of 100% of allelic diversity, we used the traditional irreplaceability measure reported by Marxan, which ranges between 100% (highly irreplaceable) and 0% (not irreplaceable). This measure is estimated by running the optimization algorithm a number of times (N=100 runs in our case) and then computing the frequency of selection of each site within the solutions obtained. Sites with unique allelic composition will be selected across all runs and reported as highly irreplaceable, whereas sites with more common alleles, replaceable by other sites with the same alleles, will appear poorly irreplaceable. For the other conservation targets, we selected a random pull of 50, 75 and 90% of the

total number of alleles existing at the basin level and for each species. We then ran the optimization algorithm to identify the minimum set of sites representing this particular selection of alleles. Given that the analyses were run only on a subset of alleles, we replicated this process 100 times to minimize the effect of the arbitrary selection of alleles. For each subset of alleles, we ran the Marxan procedure as explained above (e.g. constant cost and 100 runs for each) and retained the best solution for subsequent analyses (solution with the lowest value for the objective function). We then calculated the irreplaceability as the frequency of selection of each site within the 100 random pull of alleles. For a given species and a given conservation target, we considered a site as irreplaceable for genetic diversity conservation when it displayed an irreplaceability value of 100%. We selected such a high threshold so as to be conservative. We tested and compared visually how the proportions of irreplaceable sites vary among species and conservation targets.

To test how pooling data from several species affect the identification of irreplaceable sites, we further performed a "pooled" analysis, in which all alleles found for each common species at a site were pooled together in a single input dataset. We then selected a random pull of 30, 50, 75, 90 and 100% of the total number of alleles existing at the basin level (all common species confounded), and performed 100 Marxan runs *per* conservation targets to identify the minimum set of sites representing these particular selections of alleles.

<u>A real conservation-oriented study using SCP approaches</u>

In this second step, we (i) explored the spatial distribution of irreplaceable sites, (ii) tested whether conservation solutions found for each species are congruent among

species, (iii) tested whether solutions found for one species can be used as a surrogate for other species, and (iv) tested whether –or not– irreplaceable sites are correctly predicted by classical indices of genetic diversity. In this step, we included the two rare species, since congruency and surrogacy are particularly important to measure for rare species for which data are more difficult to collect. We therefore focus more specifically on the comparisons implying common *vs*. rare species.

Identification of irreplaceable sites for genetic conservation. We focused only on irreplaceable sites identified for the 90% target and used the program Marxan as described above to identify these sites for *L. burdigalensis* and *P. toxostoma* independently, in addition of the four common species.

We mapped these irreplaceable sites (for the six species pooled or independently) on the river network, so as to test (i) whether or not specific areas harboured more irreplaceable sites (e.g. upstream areas that are generally thought to be of high conservation priority[29]) and (ii) if irreplaceable sites are spatially congruent among species and, most notably, among common and rare species. In addition, we ran GLMs (assuming a binomial error terms distribution) including whether or not a site has been designated as an irreplaceable site at the 90% target as a binomial dependent variable, and distance to the outlet of sampling sites and betweenness centrality values[49,50] for each sites explanatory variables. Betweenness centrality is an index quantifying the positional importance of each sampling site within the river basin[18,32]. The significance of each term was tested at the α=0.05 threshold.

Surrogacy in irreplaceable sites among species. We then estimated the levels of surrogacy among species by calculating the percentage of the total number of alleles

observed for a given species that is covered by irreplaceable sites identified for another species. Although surrogacy was calculated for all species pairs, we specifically focused on rare species by calculating (i) the percentage of the total number of alleles observed for rare species covered by all the irreplaceable sites identified for all the common species, and (ii) the percentage of the total number of alleles observed for common species covered by all the irreplaceable sites identified for the rare species.

Relationships between irreplaceable sites and indices of genetic diversity. To test the ability of classical genetic indices to predict the propensity for a site to be irreplaceable from a conservation viewpoint, we ran for each species Generalized Linear Models (GLMs, with a binomial error terms distribution) including whether or not a site has been designated as an irreplaceable site at the 90% target as a binomial dependent variable, and *AR*, *PA* and D_{est} as explanatory variables. We tested the significance of each term at the α=0.05 threshold. Explanatory variables were centred and scaled, in order to compare the relative strength of the predictors among species.

Acknowledgements

We are very grateful to all the ONEMA staff that contributed to the field sampling, in particular Laurence Blanc. Leslie Faggiano, Charlotte Evangelista, Christine Lauzeral, Loïc Tudesque, Roselyne Étienne, Sébastien Villéger, Julien Cucherousset, Vincent Dubut, Alain Blanchet and South-Western France ONEMA staff are thanked for their help in the field. Roselyne Étienne and Vincent Dubut are also thanked for their help with molecular analyses. Lisa Fourtune is acknowledged for having calculated the betweenness centrality values of our river network. We also thank

Marie-Hélène Lizée for her help with GIS tools and Camille Pagès for her comments. Further, the authors thank the "Génopole Toulouse" for help with genotyping and Keoni Saint-Pée for correcting the English. IPV was financially supported by a MESR ("Ministère de l'Enseignement Supérieur et de la Recherche") PhD scholarship during this study. VH was funded by the National Environmental Research Program Northern Australia Hub and a "Ramon y Cajal" contract (RYC-2013-13979) funded by the Spanish government. This study is part of the European project "IMPACT". This project has been carried out with financial support from the Commission of the European Communities, specific RTD programme "IWRMNET". This work has been done in two research units (EDB & SETE) that are part of the "Laboratoire d'Excellence" (LABEX) entitled TULIP (ANR-10-LABX-41).

References

1. Watson, J. E. M., Dudley, N., Segan, D. B. & Hockings, M. The performance and potential of protected areas. *Nature* **515,** 67–73 (2014).
2. Filipe, A. F. *et al.* Selection of priority areas for fish conservation in Guadiana River Basin, Iberian Peninsula. *Conserv. Biol.* **18,** 189–200 (2004).
3. Kirkpatrick, J. B. An iterative method for establishing priorities for the selection of nature reserves: An example from Tasmania. *Biol. Conserv.* **25,** 127–134 (1983).
4. Margules, C. R. & Pressey, R. L. Systematic conservation planning. *Nature* **405,** 243–253 (2000).
5. Asmyhr, M. G., Linke, S., Hose, G. & Nipperess, D. A. Systematic conservation planning for groundwater ecosystems using phylogenetic diversity. *PLoS ONE* **9,** e115132 (2014).

6. Buerki, S. *et al.* Incorporating evolutionary history into conservation planning in biodiversity hotspots. *Philos. Trans. R. Soc. B Biol. Sci.* **370,** 20140014–20140014 (2015).
7. Thomassen, H. A. *et al.* Mapping evolutionary process: a multi-taxa approach to conservation prioritization. *Evol. Appl.* **4,** 397–413 (2011).
8. Diniz-Filho, J. A. F. *et al.* Planning for optimal conservation of geographical genetic variability within species. *Conserv. Genet.* **13,** 1085–1093 (2012).
9. Hermoso, V. *et al.* Species distributions represent intraspecific genetic diversity of freshwater fish in conservation assessments. *Freshw. Biol.* **61,** 1707–1719 (2016).
10. Hoffmann, A. A. & Sgrò, C. M. Climate change and evolutionary adaptation. *Nature* **470,** 479–485 (2011).
11. Carroll, S. P. *et al.* Applying evolutionary biology to address global challenges. *Science* **346,** 1245993–1245993 (2014).
12. Mittell, E. A., Nakagawa, S. & Hadfield, J. D. Are molecular markers useful predictors of adaptive potential? *Ecol. Lett.* **18,** 772–778 (2015).
13. Rey, O., Danchin, E., Mirouze, M., Loot, C. & Blanchet, S. Adaptation to global change: A transposable element–epigenetics perspective. *Trends Ecol. Evol.* **31,** 514–526 (2016).
14. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155,** 945–959 (2000).
15. Grantham, H. S., Pressey, R. L., Wells, J. A. & Beattie, A. J. Effectiveness of biodiversity surrogates for conservation planning: Different measures of effectiveness generate a kaleidoscope of variation. *PLoS ONE* **5,** e11430 (2010).

16. Hermoso, V., Januchowski-Hartley, S. R. & Pressey, R. L. When the suit does not fit biodiversity: Loose surrogates compromise the achievement of conservation goals. *Biol. Conserv.* **159,** 197–205 (2013).
17. Taberlet, P. *et al.* Genetic diversity in widespread species is not congruent with species richness in alpine plant communities. *Ecol. Lett.* **15,** 1439–1448 (2012).
18. Fourtune, L., Paz-Vinas, I., Loot, G., Prunier, J. G. & Blanchet, S. Lessons from the fish: a multi-species analysis reveals common processes underlying similar species-genetic diversity correlations. *Freshw. Biol.* **61,** 1830–1845 (2016).
19. Seymour, M., Seppälä, K., Mächler, E. & Altermatt, F. Lessons from the macroinvertebrates: species-genetic diversity correlations highlight important dissimilar relationships. *Freshw. Biol.* **61,** 1819–1829 (2016).
20. Andrew, R. L. *et al.* A road map for molecular ecology. *Mol. Ecol.* **22,** 2605–2626 (2013).
21. Pauls, S. U. *et al.* Integrating molecular tools into freshwater ecology: developments and opportunities. *Freshw. Biol.* **59,** 1559–1576 (2014).
22. Frankham, R., Briscoe, D. A. & Ballou, J. D. *Introduction to conservation genetics*. (Cambridge University Press, 2002).
23. Neel, M. C. & Cummings, M. P. Effectiveness of conservation targets in capturing genetic diversity. *Conserv. Biol.* **17,** 219–229 (2003).
24. Jost, L. G(ST) and its relatives do not measure differentiation. *Mol. Ecol.* **17,** 4015–4026 (2008).
25. Peterson, E. & Ver Hoef, J. M. V. STARS: An ArcGIS toolset used to calculate the spatial information needed to fit spatial statistical models to stream network data. *J. Stat. Softw.* **56,** 1–17 (2014).

26. Ver Hoef, J., Peterson, E., Clifford, D. & Shah, R. SSN: An R package for spatial statistical modeling on stream networks. *J. Stat. Softw.* **56,** 1–43 (2014).

27. Fortuna, M. A., Albaladejo, R. G., Fernandez, L., Aparicio, A. & Bascompte, J. Networks of spatial genetic variation across species. *Proc. Natl. Acad. Sci.* **106,** 19044–19049 (2009).

28. Paz-Vinas, I., Loot, G., Stevens, V. M. & Blanchet, S. Evolutionary processes driving spatial patterns of intraspecific genetic diversity in river ecosystems. *Mol. Ecol.* **24,** 4586–4604 (2015).

29. Finn, D. S., Bonada, N., Múrria, C. & Hughes, J. M. Small but mighty: headwaters are vital to stream network biodiversity at two levels of organization. *J. North Am. Benthol. Soc.* **30,** 963–980 (2011).

30. Paz-Vinas, I. & Blanchet, S. Dendritic connectivity shapes spatial patterns of genetic diversity: a simulation-based study. *J. Evol. Biol.* **28,** 986–994 (2015).

31. Thomaz, A. T., Christie, M. R. & Knowles, L. L. The architecture of river networks can drive the evolutionary dynamics of aquatic populations. *Evolution* **70,** 731–739 (2016).

32. Altermatt, F. Diversity in riverine metacommunities: a network perspective. *Aquat. Ecol.* **47,** 365–377 (2013).

33. Mimura, M. *et al.* Understanding and monitoring the consequences of human impacts on intraspecific variation. *Evol. Appl.* **10,** 121–139 (2017).

34. Siefert, A. *et al.* A global meta-analysis of the relative extent of intraspecific trait variation in plant communities. *Ecol. Lett.* **18,** 1406–1419 (2015).

35. Kottelat, M. & Freyhof, J. *Handbook of European freshwater fishes*. (Publications Kottelat, 2007).

36. Poulet, N., Beaulaton, L. & Dembski, S. Time trends in fish populations in metropolitan France: insights from national monitoring data. *J. Fish Biol.* **79,** 1436–1452 (2011).

37. Paz-Vinas, I., Quéméré, E., Chikhi, L., Loot, G. & Blanchet, S. The demographic history of populations experiencing asymmetric gene flow: combining simulated and empirical data. *Mol. Ecol.* **22,** 3279–3291 (2013).

38. Paz-Vinas, I. *et al.* Combining genetic and demographic data for prioritizing conservation actions: insights from a threatened fish species. *Ecol. Evol.* **3,** 2696–2710 (2013).

39. Aljanabi, S. M. & Martinez, I. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Res.* **25,** 4692–4693 (1997).

40. Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. & Shipley, P. micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **4,** 535–538 (2004).

41. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24,** 1403–1405 (2008).

42. Rousset, F. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* **8,** 103–106 (2008).

43. Szpiech, Z. A., Jakobsson, M. & Rosenberg, N. A. ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* **24,** 2498–2504 (2008).

44. Petit, R. J., El Mousadik, A. & Pons, O. Identifying populations for conservation on the basis of genetic markers. *Conserv. Biol.* **12,** 844–855 (1998).

45. Kalinowski, S. T. Counting alleles with rarefaction: Private alleles and hierarchical sampling designs. *Conserv. Genet.* **5,** 539–543 (2004).

46. Winter, D. J. mmod: an R library for the calculation of population differentiation statistics. *Mol. Ecol. Resour.* **12,** 1158–1160 (2012).

47. Ball, I. R., Possingham, H. P. & Watts, M. in *Spatial conservation prioritisation: Quantitative methods and computational tools* (eds. Moilanen, A., Wilson, K. A. & Possingham, H. P.) **Chapter 14,** 185–195 (Oxford Iniversity Press, Oxford, UK, 2009).

48. Hermoso, V. & Kennard, M. J. Uncertainty in coarse conservation assessments hinders the efficient achievement of conservation goals. *Biol. Conserv.* **147,** 52–59 (2012).

49. Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry* **40,** 35 (1977).

50. Estrada, E. & Bodin, Ö. Using network centrality measures to manage landscape connectivity. *Ecol. Appl.* **18,** 1810–1825 (2008).

Author Contributions: The study was designed by SB, GL and IPV. IPV and SB wrote the manuscript with the help of VH, GG, NP and GL. IPV, SB, GL and CV collected the samples. GL and CV produced the genetic data. IPV, VH and SB conducted the population genetic and statistical analyses.

Competing Financial Interests statement: Authors declare not having any competing financial interest.

TABLE 1: Table reporting for each species the percentages of their total number of alleles that are covered by the sampling sites having been identified as irreplaceable sites for the other species (considering the 90% conservation target).

Proportion of alleles from (below) that fell within irreplaceable sites from (right):	*Squalius cephalus*	*Gobio occitaniae*	*Barbatula barbatula*	*Phoxinus phoxinus*	*Leuciscus burdigalensis*	*Parachondrostoma toxostoma*	Common species	Rare species
Squalius cephalus	*90.00*	17.66	61.52	56.52	61.60	39.73	—	65.18
Gobio occitaniae	65.60	*90.00*	66.76	79.58	70.27	50.97	—	73.40
Barbatula barbatula	49.79	31.65	*90.00*	79.57	76.78	38.95	—	79.40
Phoxinus phoxinus	48.25	42.60	68.43	*90.00*	63.49	44.61	—	68.88
Leuciscus burdigalensis	44.73	0	66.13	67.09	*90.00*	36.42	79.87	—
Parachondrostoma toxostoma	78.55	0	78.55	50.00	67.14	*90.00*	90.00	—
Average over all other species	*55.58*	*18.38*	*68.28*	*66.55*	*67.86*	*42.14*		

Figures legends

Figure 1

Spatial distribution of observed (coloured circles) and interpolated (coloured lines) values of *AR*, *PA* and D_{est} (A, B, C respectively in the figure) for *Squalius cephalus*, *Gobio occitaniae*, *Barbatula barbatula*, and *Phoxinus phoxinus* (1, 2, 3, 4 respectively in the figure) obtained with GLMSSNs. The width of coloured lines is inversely related to the prediction standard error. The cursor on the vertical coloured scale indicates the mean value of *AR*, *PA* and D_{est}. The slope (β) of each explanatory variable (i.e. topological distance to the outlet, longitude and latitude) and its significance is also reported. N.I. indicates that the explanatory variable has not been included in the model; * p-value <0.05; ** p-value <0.01; *** p-value <0.001.

Figure 2

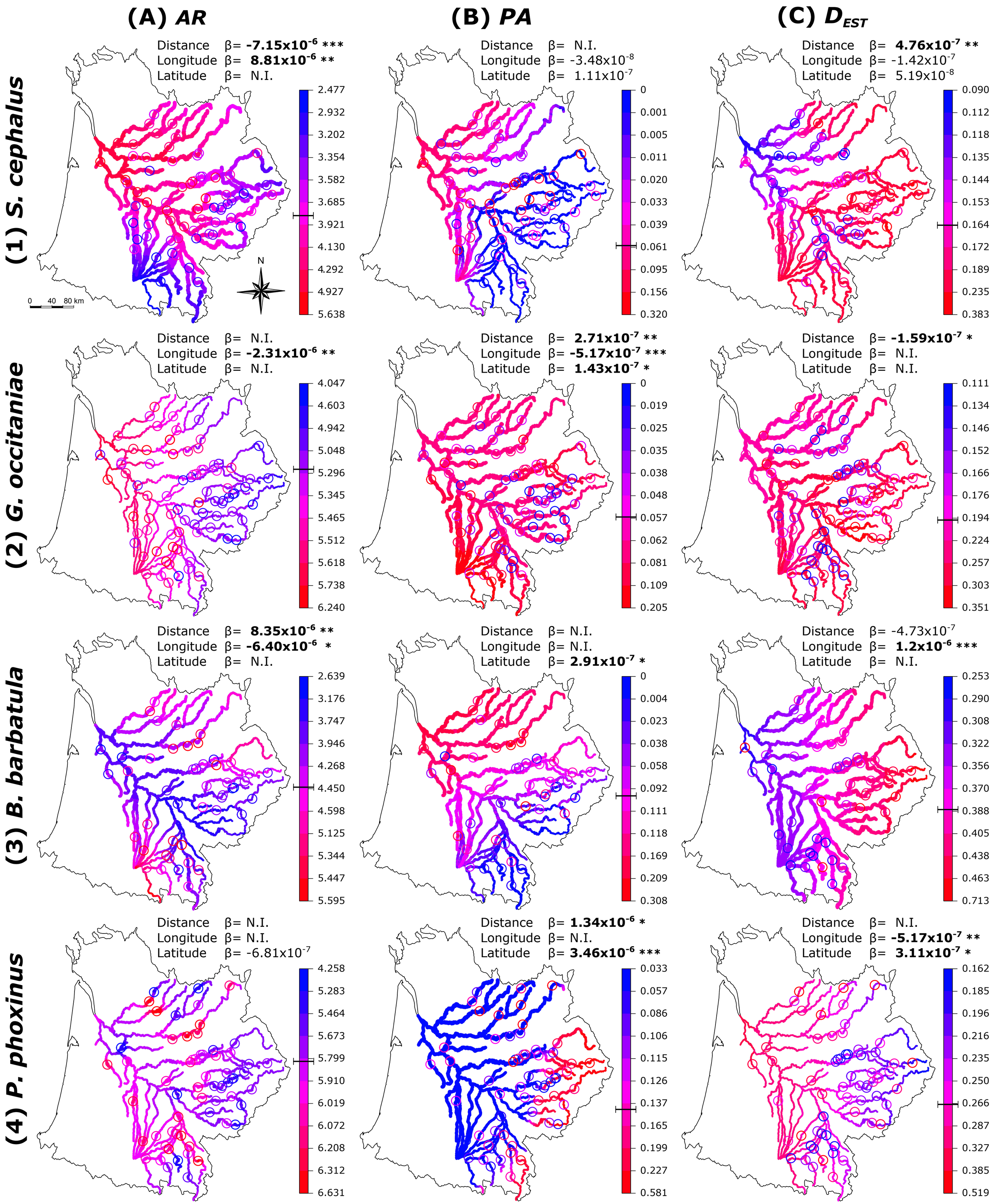
Percentage of irreplaceable sites identified by Marxan for conservation targets of 50, 75, 90 and 100% of the total number of alleles present in the river basin for each common species and for a pooled analysis in which all alleles from all common species were pooled together. N_{SITES} represents the number of sites included in the Marxan analyses.

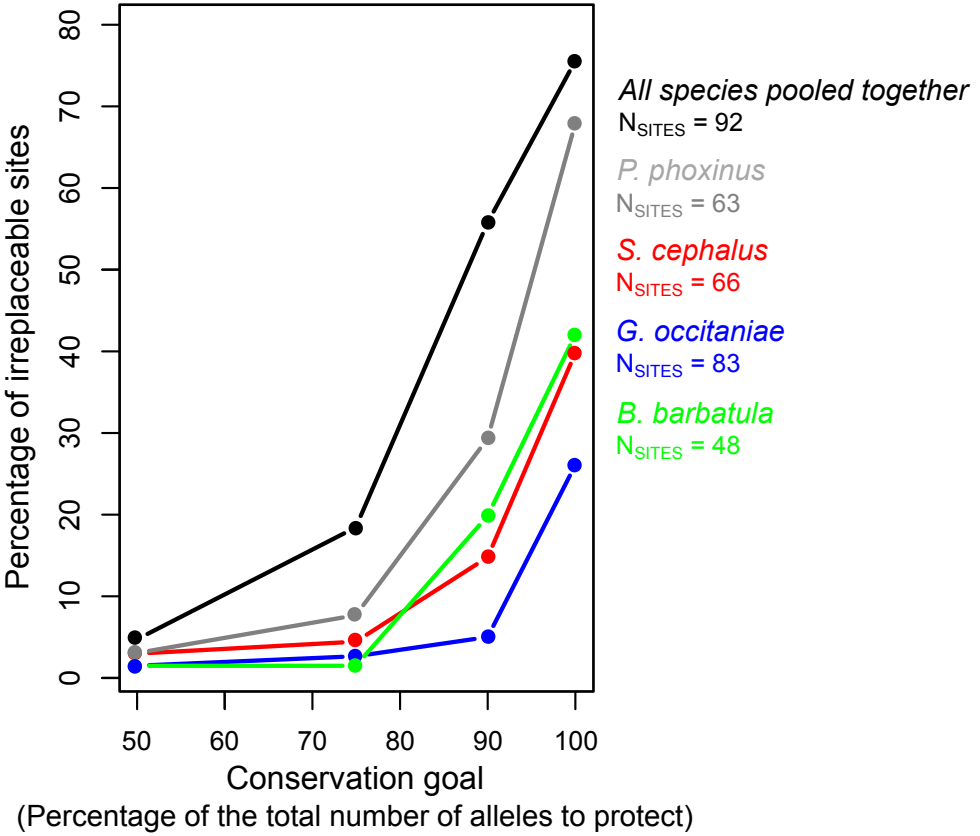
Figure 3

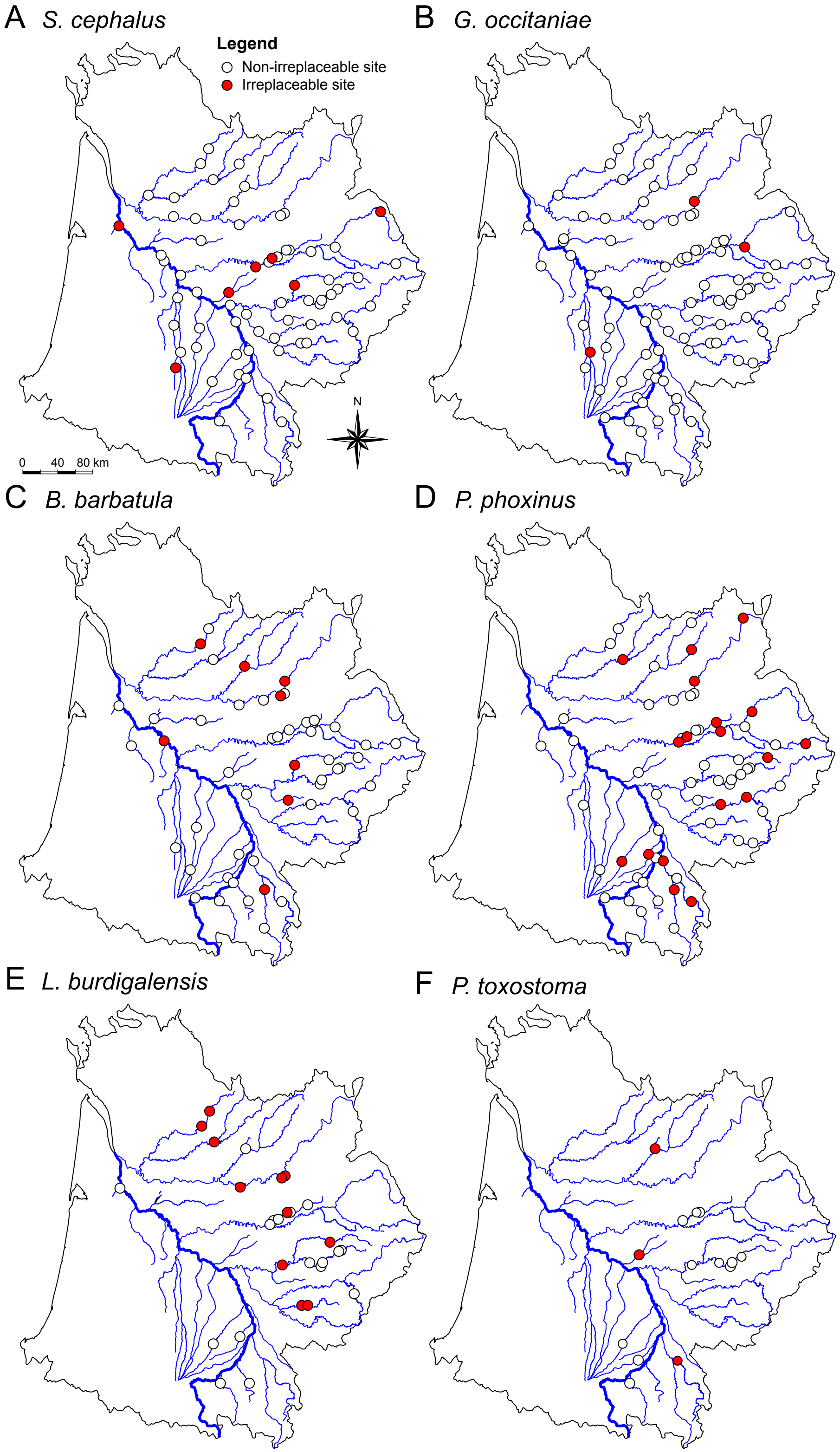
Irreplaceable sites that have been identified for each species by Marxan for preserving 90% of the total number of alleles present in the river basin (red-filled circles).

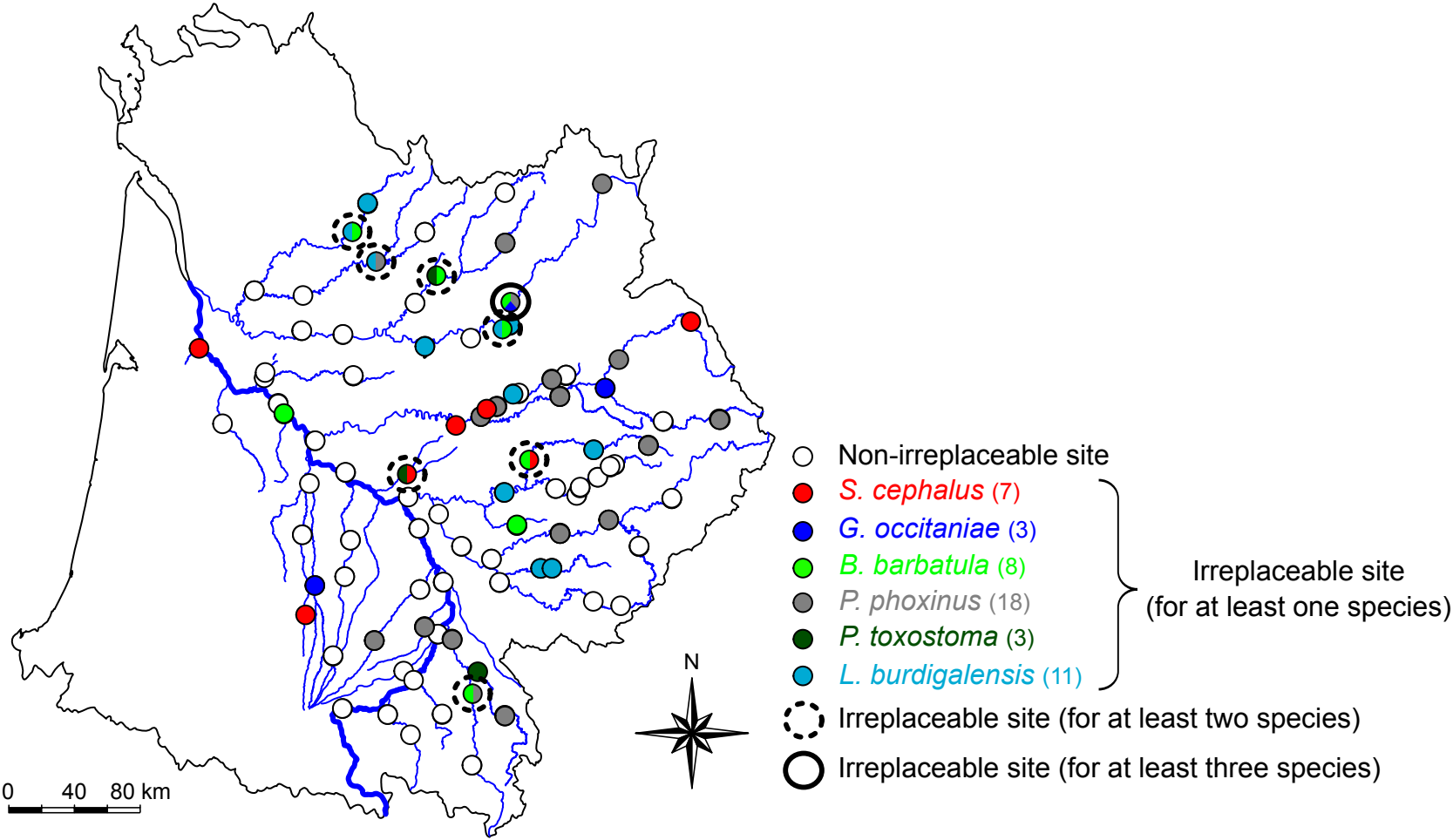
Figure 4

Irreplaceable sites that have been identified by Marxan for at least one (unicoloured-filled points), two (black dotted circles surrounding bicoloured points) or three (black bolded circle surrounding a tricoloured point) species, assuming a conservation target of 90% of the total number of alleles present in the river basin when considering the six species.


(A) AR
(B) PA
(C) DEST
(1) S. cephalus
(2) G. occitaniae
(3) B. barbatula
(4) P. phoxinus
Distance β= -7.15x10-6 ***
Longitude β= 8.81x10-6 **
Latitude β= N.I.
2.477 2.932 3.202 3.354 3.582 3.685 3.921 4.130 4.292 4.927 5.638
Distance β= N.I.
Longitude β= -3.48x10-8
Latitude β= 1.11x10-7
0 0.001 0.005 0.011 0.020 0.033 0.039 0.061 0.095 0.156 0.320
Distance β= 4.76x10-7 **
Longitude β= -1.42x10-7
Latitude β= 5.19x10-8
0.090 0.112 0.118 0.135 0.144 0.153 0.164 0.172 0.189 0.235 0.383
Distance β= N.I.
Longitude β= -2.31x10-6 **
Latitude β= N.I.
4.047 4.603 4.942 5.048 5.296 5.345 5.465 5.512 5.618 5.738 6.240
Distance β= 2.71x10-7 **
Longitude β= -5.17x10-7 ***
Latitude β= 1.43x10-7 *
0 0.019 0.025 0.035 0.038 0.048 0.057 0.062 0.081 0.109 0.205
Distance β= -1.59x10-7 *
Longitude β= N.I.
Latitude β= N.I.
0.111 0.134 0.146 0.152 0.166 0.176 0.194 0.224 0.257 0.303 0.351
Distance β= 8.35x10-6 **
Longitude β= -6.40x10-6 *
Latitude β= N.I.
2.639 3.176 3.747 3.946 4.268 4.450 4.598 5.125 5.344 5.447 5.595
Distance β= N.I.
Longitude β= N.I.
Latitude β= 2.91x10-7 *
0 0.004 0.023 0.038 0.058 0.092 0.111 0.118 0.169 0.209 0.308
Distance β= -4.73x10-7
Longitude β= 1.2x10-6 ***
Latitude β= N.I.
0.253 0.290 0.308 0.322 0.356 0.370 0.388 0.405 0.438 0.463 0.713
Distance β= N.I.
Longitude β= N.I.
Latitude β= -6.81x10-7
4.258 5.283 5.464 5.673 5.799 5.910 6.019 6.072 6.208 6.312 6.631
Distance β= 1.34x10-6 *
Longitude β= N.I.
Latitude β= 3.46x10-6 ***
0.033 0.057 0.086 0.106 0.115 0.126 0.137 0.165 0.199 0.227 0.581
Distance β= N.I.
Longitude β= -5.17x10-7 **
Latitude β= 3.11x10-7 *
0.162 0.185 0.196 0.216 0.235 0.250 0.266 0.287 0.327 0.385 0.519
N
0 40 80 km


Percentage of irreplaceable sites
0
10
20
30
40
50
60
70
80
50
60
70
80
90
100
Conservation goal
(Percentage of the total number of alleles to protect)
All species pooled together
N SITES = 92
P. phoxinus
N SITES = 63
S. cephalus
N SITES = 66
G. occitaniae
N SITES = 83
B. barbatula
N SITES = 48


A S. cephalus
Legend
Non-irreplaceable site
Irreplaceable site
N
0 40 80 km
B G. occitaniae
C B. barbatula
D P. phoxinus
E L. burdigalensis
F P. toxostoma


Non-irreplaceable site
S. cephalus (7)
G. occitaniae (3)
B. barbatula (8)
P. phoxinus (18)
P. toxostoma (3)
L. burdigalensis (11)
Irreplaceable site
(for at least one species)
Irreplaceable site (for at least two species)
Irreplaceable site (for at least three species)
N
0 40 80 km